An adaptive vocal-user-interface learning from speech examples of the user

Bart Ons, Jort F. Gemmeke and Hugo Van hamme

Department ESAT-PSI, KULeuven, Leuven, Belgium

In the ALADIN (Adaptation and Learning for Assistive Domestic Vocal Interfaces) project, we aim to develop a self-learning vocal user interface (VUI) that learns from interaction with its users, including persons with a speech impairment. The challenge of the project is to equip the VUI with speech and machine learning algorithms to make the VUI adaptive to the user. Here we present the technical status of the project.

The VUI training process is grounded in the environment of the user by mining the speech input from the user and the actions provoked on a device. Two modules play a central role in training and decoding spoken commands: the ***word finding module*** and the ***grammar induction module*** [1]. The word finding module looks for word-sized recurring acoustic patterns in the audio input that correlate well with the provoked action. The description of the provoked action serves as weak supervision in searching recurrent acoustic patterns and strengthen the discovery of word-sized units in the acoustic signal. The word finding module is based on weakly supervised NMF (non-negative matrix factorization) [2]. The description of the provoked action and the output of the word finding module serve as input to the grammar induction module which learns the final mapping between the word order and the provoked action. The grammar induction technique is based on semi-supervised hidden Markov model (HMM) learning.

We investigated whether enhanced acoustic representations allow for faster learning in the word finding module. We found that representations building on prior knowledge like HMM-based phone models accelerate word learning and user-dependent acoustic representations allow for better adaptability [3]. We also investigated whether we could gain performance by grammar induction and found that grammar induction led to significant gains [4].

References

[1] Gemmeke J.F., van de Loo J., De Pauw G., Driesen J., Van hamme H. and Daelemans W (2012). A Self-Learning Assistive Vocal Interface Based on Vocabulary Learning and Grammar Induction. In *Proc. INTERSPEECH*, 2012.

[2] Driesen J., Gemmeke J.F., and Van hamme H. (2012) Weakly supervised keyword learning using sparse representations of speech. *in Proceedings of the 36th International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, 2012*.

[3] Ons,B., Gemmeke, J.F., and Van hamme, H. (2013). *Fast Word learning in a self-learning vocal user interface*. Submitted.

[4] Ons,B., Tessema, N., van de Loo, J., Gemmeke, J.F., De Pauw, G., Daelemans, W., and Van hamme, H. (2013). *A Self Learning Vocal Interface for Speech-impaired Users*. Submitted.